Location Obfuscation using Smart Meter Readings

Kiran Maharaj and Patrick Hosein
Department of Computer Science
The University of the West Indies, St. Augustine, Trinidad
Email: kiran.maharaj4@my.uwi.edu, patrick.hosein@sta.uwi.edu

***Abstract*—The World Wide Web has become a popular platform for the hosting of surveys. Such online surveys can more easily and efficiently collect data and are accessible using both desktop and mobile devices. In many cases personal information, such as postal addresses, is requested by these surveys. With the availability of location detection technology in mobile devices, some surveys even go further and request spatial coordinates (e.g. GPS information). This raises the issue of location privacy since this may lead to privacy attacks. One can prevent this by providing vague location information but, in many cases, this may defeat the purpose of having the location field since it may be required by the surveyors. Hence there is a delicate balance between user privacy and location accuracy. In this paper we address this trade-off by proposing a framework for location obfuscation in online surveys. In this framework we allow users to state their privacy requirement using a target probability which is the probability of they being correctly identified. We then provide the most accurate location information to the surveyors given the desired privacy level. This approach requires population density information and so we also provide an interesting approach to estimating this information from mining Smart Meter data.**

***Index Terms*—location obfuscation, location privacy, online survey, smart meters**

I. Introduction

With the increasing popularity of the Internet and the easy availability of smart phones, online surveys have become a common method for individuals and companies to gather opinions and information on products and social topics. Such surveys are designed to capture a wide range of information and typically includes demographic information about the survey population. Traditional surveys do this by asking personal questions such as name, age, ethnicity and/or postal address of the person. However, anonymous surveys are sometimes used to protect the identity of the respondent. In such cases no personal information is requested. However, if the surveyor requires location information then one must determine how this can be provided while maintaining privacy. For example, the respondent can be asked to provide less specific information such as their street name or city or Postal Code in order to protect their location privacy. This information however may end up being too specific (e.g., if they live on a very short street) or provide too little location information for the purposes of the survey (e.g., if they live on a very long street). With the advent of smart phones one can instead request GPS coordinates either automatically (through a mobile application) or manually. Such submissions also have the advantage of easy presentation on maps and are now available in online survey tools such as Snap Surveys and Survey Swipe. However, many users would prefer not to provide such specific information and hence some mechanism must be provided for them to provide less specific location information (as they would have done in a traditional survey). For example, a nearby location can instead be provided that would make it difficult to determine the respondent. This paper focuses on providing such a mechanism.

Location Privacy allows an individual to determine when, how and to what extent their location information should be communicated to others. Given a person's location one can obtain personal information about the person such as name, age and political interests [1], [2]. Privacy invasion attacks continue to grow as technology makes it more easily possible to provide and obtain precise user locations. The increased interest in location privacy has led to a surge of research into countermeasures and protection methods against potential threats. These methods fall into non-computational countermeasures (e.g., regulatory strategies and privacy policies) and computational countermeasures (e.g., anonymity and obfuscation methods as outlined in surveys by Duckham and Kulick[3] and Wernke et al. [4]). We focus on computational measures as these would be more applicable for online surveys. Location Obfuscation can be defined as a set of techniques used to degrade the quality of location information by a specified level. The essential idea is to alter the original location of a user to ensure the person's original location cannot be easily determined. The aim in such alteration is to allow a user to provide a location they are comfortable with to some third-party. Location Obfuscation methods are complimentary to Anonymity methods [5] which focus on the disassociation of location information and a persons identity. We thus consider and adapt some Anonymity Methods in our work but focus on the perturbation of location information for use in surveys.

Many developed countries support an Automatic Meter Reading (AMR) network in which smart meters are placed at users' homes and electrical usage readings are transmitted wirelessly and collected for billing purposes. The GPS coordinates of such smart meters are also known. Furthermore, assuming that almost all of the population has electricity (which holds in developed countries) then one can safely assume that the location of such meters correlates well with population density. In this paper, we propose a framework to perform Location Obfuscation in Online Surveys using

978-1-5090-2576-3/16/$31.00 ©2016 IEEE

Smart Meter information. We first discuss various location obfuscation techniques and then propose a modification of the K-Anonymity Method first proposed in [6], [7]. In this method we define the privacy preference as a target probability that is specified by the user. This target probability is defined as the probability of correctly identifying the location of the target user by someone who has no knowledge of the obfuscation method. We illustrate the performance of this method through numerical results.

II. Related work

Much of the work on location privacy has been focused on Location Based Services for Mobile Networks [8], [9], [6], [10], [7]. Such work highlights the variety of computational countermeasures and strategies proposed to protect the location privacy of individuals. Approaches by the authors highlight differences in user expectations, the countermeasure methodology as well as how privacy can be measured [11]. In this section we outline several proposed methods and highlight why they are not suitable for our needs.

A. Obfuscation Methods

In selecting obfuscation techniques we must consider the objective of obfuscating locations in an online survey. Therefore (a) the technique must be applicable to single sample geographic locations and likewise only result in the return of a single obfuscated location and (b) the technique must use a comprehensible privacy preference which can be adjusted by users. Given these criteria we consider four obfuscation techniques.

1) Decimal Rounding: This is a manual method of obfuscation which can be done by the user or through a simple rounding technique built into the survey application. The technique involves the rounding of a geographic location to n decimal places. A simplification of the rounding algorithm proposed in [12], this technique provides a high level of obfuscation when n is small. Note that reducing the number of significant decimal places provides very coarse control. However we propose more granular control by doing the following. Consider some integer K. Let Y denote the exact value of the coordinate (either longitude or latitude). We obtain the obfuscated value of this coordinate as

$$X = \frac{\lfloor KY + 0.5 \rfloor}{K} \tag{1}$$

For example if $K = 10$ then Y is rounded to one decimal place while if $K = 100$ it is rounded to two decimal places. However, with this approach one can use values of K between 10 and 100 and these would provide accuracy levels between those obtained with $K = 10$ and $K = 100$. In general, larger values of K provide more accuracy.

2) Privacy Radius: A family of point-based location obfuscation techniques is introduced and evaluated by Wightman et al. in [13]. These techniques are based on the concept of producing an obfuscated value that lies within a circle of radius r centered at the original location [5], [14]. However, note that such techniques are agnostic to the population density in the area. Hence a radius that works well in a city area may not work well in a sparsely populated area since, in this case, few people will lie within the circle. Also of note is the fact that a population would not be uniformly distributed across an area, these obfuscation techniques would thus be unable to ensure the probability of attack success across a chosen circle.

3) K-Anonymity: The K-anonymity concept follows the idea that a person is K-anonymous if he/she cannot be distinguished from $K-1$ other people. Originally proposed as a method for protecting location privacy by Gruteser and Grunwald [6], this method involves the determination of an area or region containing K people including the target user. The computation of this area was originally done using quadtrees however a number of variations have been proposed by Duckham and Kulick [15] using a graph based approach. Another method is that of Gedik and Liu [7], [16] who introduced the idea of allowing end-user adjustments of K which allows users to specify their level of anonymity. The K-anonymous approach does have drawbacks since it requires generation of an area [17] as opposed to a single point which is required in our case. The proposed obfuscation method in this paper selects a single person and their location within a set of at least K people. The probability of successfully determining the user is therefore at most $1/K$ and hence the user is at least K-anonymous.

4) Spatial Discretization: In this approach a grid is formed over the area that includes the target user. The obfuscated location is obtained as the closest grid vertex to the target user. Note that, in this case, many locations are mapped to the same obfuscated value. As with the privacy radius approach, the size of the grid spacing determines how well the method performs. A grid spacing that works well in a city area may not work well in a rural area. Hence the grid spacing will have to be varied based on population density and this process can be quite complicated.

B. Discussion

Note that, for our work, we have two objectives user privacy and location accuracy. In highly populated areas we can provide good location accuracy and user privacy since, if the obfuscated point is within a small area around the target location there are still many people within which to ”hide” the user. However this is not the case for rural areas where less accuracy must be provided in order to guarantee the same level of privacy. This means that our method must include population density. Hence those approaches above that are density agnostic are not suitable. However the ones that do take into account density do so based on provided population density information. We will provide an approach that uses population density but does not directly compute density information.

III. Framework

In this section we describe the framework used for our proposed solution. Key to this framework are the following assumptions (a) a smart grid network is available and the locations of the smart meters are given and (b) each household has one of these smart meters. We use historical smart meter data to determine (using Machine Learning methods) the number of people in a household and then use this together with the meter location to estimate population density. Once this computation is performed we no longer need the Smart Meter information unless we need to update our population density information. So note that no real-time access to smart meters are needed and no information of users or their meter locations is made available to the public.

A. Smart Grid Overview

Smart grids use an Advanced Metering Infrastructure (AMI) to more efficiently support the transportation, distribution and consumption of electrical energy. It is an integrated system of smart meters, communications networks and data management systems that facilitate real-time two way communication between utilities and consumers [18]. Thus, an AMI allows utilities to access an abundance of information. This information includes electrical consumption data, load profile data, demand, time-of-use, voltage profile data and power quality data [19]. Also known is the GPS coordinates of each smart meter. We use this data to estimate household size (larger households consume more energy) and only a single historical snapshot is required for this computation. Due to space limitations the details of the Machine Learning approach will not be provided.

B. Privacy Preference

In our framework we propose the use of a target probability as the privacy preference to be chosen by a user. This is the probability that the user can be identified. In practice, users may not understand such a concept and hence a more user friendly interface may be needed. For example, within the survey, a slider can be presented that goes from 1 to 1000. The user can be asked to choose a value such that, to find them, an attacker would have to make a guess among that number of people. So if they chose 100 then their privacy probability would be 0.01.

C. Obfuscation Method

For a given user's location X, we want to obtain an obfuscated value Y which provides the user with sufficient privacy while at the same time providing sufficient location information for use in a survey. For example, suppose we have a survey on the Zika virus and would like to know the extent of the spread of the virus. One can see that location information is vital since such information can be used for mosquito eradication strategies.

Let us assume that we wish to guarantee that, given a user's obfuscated location, the probability that an attacker guesses the user is no more than P. We assume that an attacker has no information other than the locations of all users. We obtain Y as follows. We find the K closest users to X, randomly choose one of these K users and provide the location of that user as Y. Suppose that there are u users at Y. Given Y the best that an attacker can do is to choose a user from that location. Their probability of success is the probability that the user is at that location, u/K, times the probability that they correctly pick the user at the location, $1/u$, which is simply $1/K$. Therefore we simply have to choose K so that $P = 1/K$.

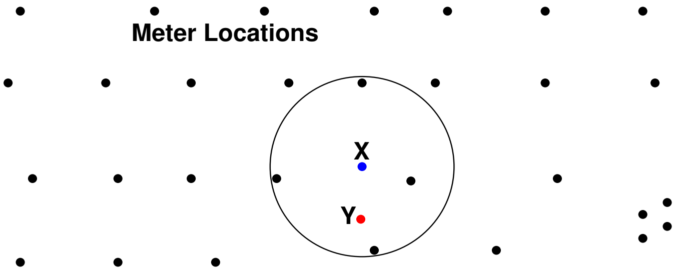

Fig. 1. Location Obfuscation Approach

We provide a simple example in Figure 1 to illustrate the approach. Here we have meter locations in an almost grid-like fashion to correspond to an urban area with rows of parallel streets. In this particular case we assume that $P = 1/18$ so that $K = 18$ and, for simplicity, we assume each household size is three. Hence we must find the five closest meters to X (which already has three users). These are shown in the circle. We then pick one of these locations Y as the obfuscated location. Note that this obfuscated value is not the same as X and thus the attacker without further information would be incorrect in determining the target.

IV. Numerical Results

In this section we provide both analytic and simulation results. In each case we are interested in two metrics, user privacy and location accuracy. We measure user privacy as follows. Given an obfuscated location, if that household is the target and it has one occupant then the attacker was successful. If the household has more than one occupant then the attacker must guess which person in the household is the target. In the case of apartment buildings, the attacker must first correctly guess the apartment location, then correctly guess the apartment within the location and finally correctly guess the person within the apartment.

The other metric is accuracy. We obtain this by measuring the location error defined as the distance between the target location and the obfuscated location. Note that, in the case of a multi-apartment building we may not care in which household the target lies since our interest is simply the location. For example, in the case of a Zika virus survey, we simply need to know the location of where the affected person lives so that appropriate actions can be taken for the area.

A. Analytic Results

In this section we consider a simple use case and analytically solve for the desired metrics. We consider a uniform grid of smart meters. Each vertex in this grid has one meter and we assume h people per household. The distance between adjacent vertices is d units. We focus on locations within the grid and so do not take into account edge effects.

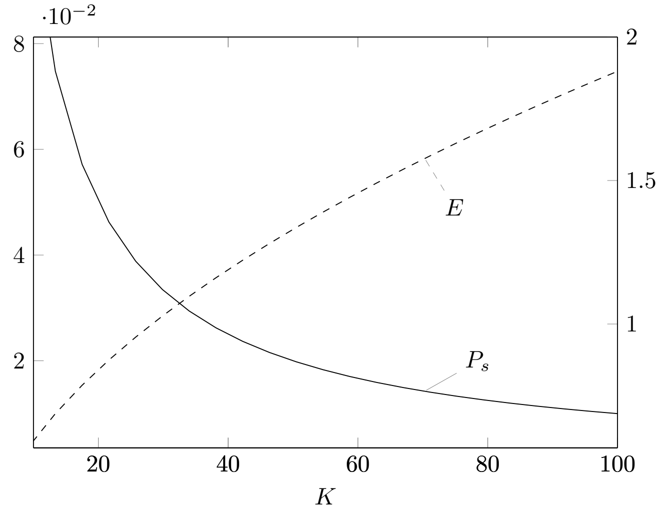

Fig. 2. Illustration of Privacy and Accuracy Trade-off as K is varied

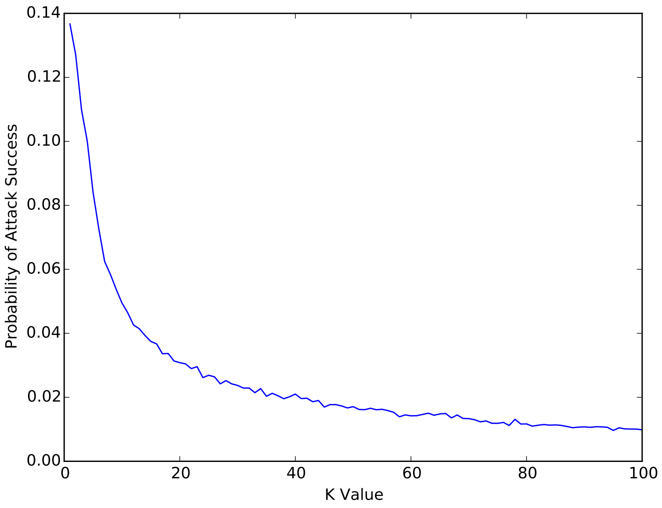

Fig. 3. Privacy Probability versus K (Urban Case)

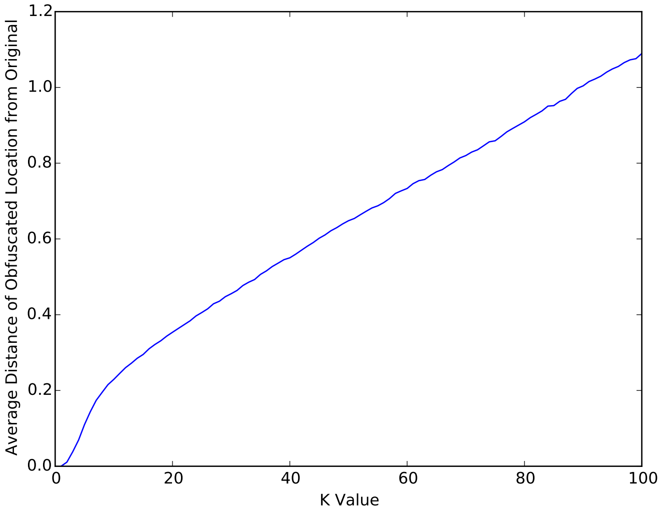

Fig. 4. Average Location Error versus K (Urban case)

Consider some target user and consider a circle around the user with radius r. The number of meters within the circle varies with r. Since we are interested in sufficiently large values of r we will obtain the number of vertices N by dividing the area of the circle with the area of N squares each of area d^2. We therefore obtain $N \approx \pi \frac{r^2}{d^2}$. With h people per household the probability of a successful attack is given by

$$P_s = \frac{1}{Nh} \approx \frac{d^2}{\pi h r^2}$$

Next we compute the expected value of the location error. This is the average distance from the origin to a point randomly dropped in a circle of radius r which is $E(r) = 2r/3$.

We provide a simple illustration of this trade-off in Figure 2. The x-axis is the number of people within the circle, $K = Nh$, the left y-axis is the probability of attacker success P_s, and the right y-axis is the accuracy, $E(r)$. For this example we used $d = 1$ and $h = 4$. As K increases the success probability decreases but the accuracy also decreases.

B. Simulation Results

In order to evaluate the proposed method, we simulate a typical urban environment. We consider ten parallel streets with ten land plots per street. We randomly choose ten of these 100 land plots and assume that they contain a multi-apartment complex containing between 2-36 apartments. We also randomly choose ten other land plots and assume that they are empty. The remaining eighty land plots are assumed to have single family homes (i.e. single meters). In total we have 200 households and hence 200 meters. We choose the size of each household using population distribution data gathered from the Australian Smart-Grid Smart-City project [20].

To simulate attack success, for each resident in the area and a given privacy probability P, we determine a set S of $\lceil \frac{1}{P} \rceil$ users closest to the target user's location. The obfuscated value is then randomly chosen from within S. The two metrics of concern are the probability of an attacker success and the average location error as previously described. In Figure 3 we plot the privacy probability (probability of attacker success) versus the number of users chosen within the set, K. In Figure 4 we plot the average error (average distance between the location and the obfuscated value) as a function of K.

From Figure 3 we note that the success probability decreases with K. Without multi-apartment buildings and empty lots one would expect this plot to follow the function $1/K$. We do see this is the case for large K. From Figure 4 we note that the location error increases almost linearly with K.

We also ran a use case for a more rural area. Here we assumed two multi-apartment buildings, 96 empty plots and 23 single family plots. In Figure 5 we plot the privacy probability while in Figure 6 we plot the average location error for each of the methods. Here we see similar results except that the average error is much larger as expected since the sufficient user samples being captured in S would be distributed in a much larger area when compared with an urban scenario.

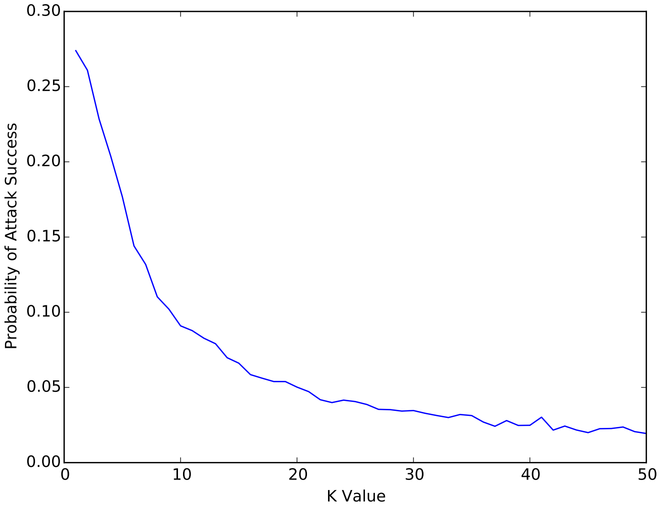

Fig. 5. Privacy Probability versus K (Rural Case)

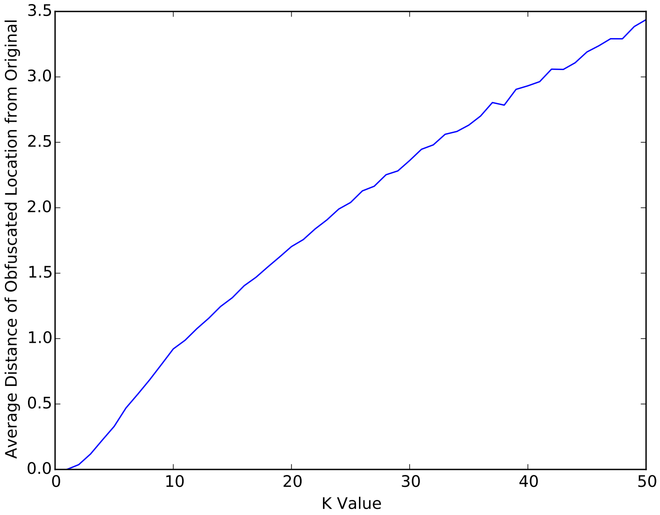

Fig. 6. Average Location Error versus K (Rural case)

V. Conclusion and Future Work

The objective of this paper was to create a framework for location obfuscation for online surveys and other applications that require a tight trade-off between location accuracy and user location privacy. We demonstrated that such an approach requires population density information since the degree of accuracy can be increased as the population density increases. The approach proposes the use of Smart Meters, used by Electricity companies, to estimate population density information. The effectiveness of the approach was demonstrated via numerical results.

We are presently planning more extensive simulations using real Smart Meter information. We are also working on Machine Learning techniques to be able to estimate household sizes. We then plan to simulate a more realistic environment. Finally we intend to deploy a prototype of the framework in order to determine its practicability.

References

[1] A. R. Beresford and F. Stajano, "Location privacy in pervasive computing," *IEEE Pervasive Computing*, vol. 2, no. 1, pp. 46–55, 2003.

[2] H. Karimi, *Advanced Location-Based Technologies and Services*. Taylor & Francis, 2013. [Online]. Available: https://books.google.com/books?id=StoLO_7cZxsC

[3] M. Duckham and L. Kulik, "Location privacy and location-aware computing," *Dynamic & mobile GIS: investigating change in space and time*, vol. 3, pp. 35–51, 2006.

[4] M. Wernke, P. Skvortsov, F. Dürr, and K. Rothermel, "A classification of location privacy attacks and approaches," in *Personal and Ubiquitous Computing*, vol. 18, no. 1, 2014, pp. 163–175.

[5] C. Ardagna, M. Cremonini, E. Damiani, S. De Capitani di Vimercati, and P. Samarati, "Location Privacy Protection Through Obfuscation-Based Techniques," *Data and Applications Security XXI*, vol. 4602, pp. 47–60, 2007. [Online]. Available: http://www.springerlink.com/content/a627753563301640

[6] M. Gruteser and D. Grunwald, "Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking," in *Proceedings of the 1st international conference on Mobile systems applications and services MobiSys 03*, no. 3, 2003, pp. 31–42. [Online]. Available: http://portal.acm.org/citation.cfm?doid=1066116.1189037

[7] B. Gedik and L. Liu, "Location Privacy In Mobile Systems A Personalized Anonymization Model," *25th IEEE International Conference on Distributed Computing Systems (ICDCS'05)*, pp. 620–629, 2005. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1437123

[8] R. Shokri, J. Freudiger, and J.-P. Hubaux, "A unified framework for location privacy," *Proceedings of the 9th International Symposium on Privacy Enhancing Technologies, PETS 2010*, pp. 203–214, 2010.

[9] J. H. Jafarian, "A vagueness-based obfuscation technique for protecting location privacy," in *Proceedings - SocialCom 2010: 2nd IEEE International Conference on Social Computing, PASSAT 2010: 2nd IEEE International Conference on Privacy, Security, Risk and Trust*, 2010, pp. 865–872.

[10] Y. Wang, D. Xu, X. He, C. Zhang, F. Li, and B. Xu, "L2P2: Location-aware location privacy protection for location-based services," in *Proceedings - IEEE INFOCOM*, 2012, pp. 1996–2004.

[11] J. Krumm, "A survey of computational location privacy," in *Personal and Ubiquitous Computing*, vol. 13, 2009, pp. 391–399.

[12] ——, "Inference Attacks on Location Tracks," *Pervasive Computing*, vol. 10, no. Pervasive, pp. 127–143, 2007. [Online]. Available: http://research.microsoft.com/en-us/um/people/jckrumm/publications 2007/inference attack refined02 distribute.pdf

[13] P. Wightman, W. Coronell, D. Jabba, M. Jimeno, and M. Labrador, "Evaluation of location obfuscation techniques for privacy in location based information systems," in *Communications (LATINCOM), 2011 IEEE Latin-American Conference on*. IEEE, 2011, pp. 1–6.

[14] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-Indistinguishability: Differential Privacy for Location-Based Systems," *Ccs'13*, vol. abs/1212.1, pp. –, 2013.

[15] M. Duckham and L. Kulik, "A formal model of obfuscation and negotiation for location privacy," *Proceedings of the 3rd International Conference on Pervasive Computing*, pp. 152–170, 2005. [Online]. Available: http://link.springer.com/chapter/10.1007/11428572_10

[16] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Transactions on Mobile Computing*, vol. 7, no. 1, pp. 1–18, 2008.

[17] L. Y. Man, C. S. Jensen, H. Xuegang, and L. Hua, "SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services," in *Proceedings - International Conference on Data Engineering*, 2008, pp. 366–375.

[18] Advanced metering infrastructure and customer systems. [Online]. Available: https://www.smartgrid.gov/recovery-act/deployment-status/ami-and-customer-systems

[19] Elster. Realizing the smart gird of the future through ami technology. [Online]. Available: http://www.elstersolutions.com/assets/downloads/WP42-1003B.pdf

[20] Department of Industry, Innovation and Science, Commonwealth of Australia. Smart-Grid Smart-City Customer Trial Data - Datasets - data.gov.au. [Online]. Available: https://data.gov.au/dataset/smart-grid-smart-city-customer-trial-data